

RAW SEQUENCE LISTING

DATE: 12/07/2001

PATENT APPLICATION: US/09/812,862

TIME: 16:35:08

Input Set : N:\Crf3\RULE60\09812862.txt

Output Set: N:\CRF3\12072001\I812862.raw

SEQUENCE LISTING

4 (1) GENERAL INFORMATION:
6 (i) APPLICANT: Wands, Jack R.
7 Scaglioni, Pier Paolo
8 Melegari, Margherita
10 (ii) TITLE OF INVENTION: INHIBITION OF HEPATITIS B REPLICATION
12 (iii) NUMBER OF SEQUENCES: 14
14 (iv) CORRESPONDENCE ADDRESS:
15 (A) ADDRESSEE: Fish & Richardson P.C.
16 (B) STREET: 225 Franklin Street
17 (C) CITY: Boston
18 (D) STATE: MA
19 (E) COUNTRY: USA
20 (F) ZIP: 02110-2804
22 (v) COMPUTER READABLE FORM:
23 (A) MEDIUM TYPE: Floppy disk
24 (B) COMPUTER: IBM PC compatible
25 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
26 (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
28 (vi) CURRENT APPLICATION DATA:
C--> 29 (A) APPLICATION NUMBER: US/09/812,862
C--> 30 (B) FILING DATE: 20-Mar-2001
31 (C) CLASSIFICATION:
37 (vii) PRIOR APPLICATION DATA:
34 (A) APPLICATION NUMBER: 08/667,073
35 (B) FILING DATE:
38 (A) APPLICATION NUMBER: 08/492,489
39 (B) FILING DATE: 20-JUN-1995
41 (viii) ATTORNEY/AGENT INFORMATION:
42 (A) NAME: Clark, Paul T.
43 (B) REGISTRATION NUMBER: 30,162
44 (C) REFERENCE/DOCKET NUMBER: 00786/282001
46 (ix) TELECOMMUNICATION INFORMATION:
47 (A) TELEPHONE: 617/542-5070
48 (B) TELEFAX: 617/542-8906
49 (C) TELEX: 200154
51 (2) INFORMATION FOR SEQ ID NO: 1:
53 (i) SEQUENCE CHARACTERISTICS:
54 (A) LENGTH: 1041 base pairs
55 (B) TYPE: nucleic acid
56 (C) STRANDEDNESS: single
57 (D) TOPOLOGY: linear
W--> 59 (ii) MOLECULE TYPE: DNA
67 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
69 ATG GAC ATA GAT CCC TAT AAA GAA TTT GGT TCA TCT TAT CAG TTG TTG 48
70 Met Asp Ile Asp Pro Tyr Lys Glu Phe Gly Ser Ser Tyr Gln Leu Leu
71 1 5 10 15

ENTERED

RAW SEQUENCE LISTING

DATE: 12/07/2001

PATENT APPLICATION: US/09/812,862

TIME: 16:35:08

Input Set : N:\Crif3\RULE60\09812862.txt

Output Set: N:\CRF3\12072001\I812862.raw

73	AAT TTT CTT CCT TTG GAC TTC TTT CCT GAC CTT AAT GCT TTG GTG GAC	96
74	Asn Phe Leu Pro Leu Asp Phe Phe Pro Asp Leu Asn Ala Leu Val Asp	
75	20 25 30	
77	ACT GCT ACT GCC TTG TAT GAA GAA GAG CTA ACA GGT AGG GAA CAT TGC	144
78	Thr Ala Thr Ala Leu Tyr Glu Glu Glu Leu Thr Gly Arg Glu His Cys	
79	35 40 45	
81	TCT CCG CAC CAT ACA GCT ATT AGA CAA GCT TTA GTA TGC TGG GAT GAA	192
82	Ser Pro His His Thr Ala Ile Arg Gln Ala Leu Val Cys Trp Asp Glu	
83	50 55 60	
85	TTA ACT AAA TTG ATA GCT TGG ATG AGC TCT AAC ATA ACT TCT GAA CAA	240
86	Leu Thr Lys Leu Ile Ala Trp Met Ser Ser Asn Ile Thr Ser Glu Gln	
87	65 70 75 80	
89	GTA AGA ACA ATC ATA GTA AAT CAT GTC AAT GAT ACC TGG GGA CTT AAG	288
90	Val Arg Thr Ile Ile Val Asn His Val Asn Asp Thr Trp Gly Leu Lys	
91	85 90 95	
93	GTG AGA CAA AGT TTA TGG TTT CAT TTG TCA TGT CTC ACT TTC GGA CAA	336
94	Val Arg Gln Ser Leu Trp Phe His Leu Ser Cys Leu Thr Phe Gly Gln	
95	100 105 110	
97	CAT ACA GTT CAA GAA TTT TTA GTA AGT TTT GTA GTA TGG ATC AGA ACT	384
98	His Thr Val Gln Glu Phe Leu Val Ser Phe Val Val Trp Ile Arg Thr	
99	115 120 125	
101	CCA GCT CCA TAT AGA CCT CCT AAT GCA CCC ATT CTC TCG ACT CTT CCG	432
102	Pro Ala Pro Tyr Arg Pro Pro Asn Ala Pro Ile Leu Ser Thr Leu Pro	
103	130 135 140	
105	GAA CAT ACA GTC ATT AGA AGA GGA GGT GCA AGA GCT TCT AGG TCC CCC	480
106	Glu His Thr Val Ile Arg Arg Gly Gly Ala Arg Ala Ser Arg Ser Pro	
107	145 150 155 160	
109	AGA AGA CGC ACT CCC TCT CCT CGC AGG AGA AGA TCC CAA AAT TCG CAG	528
110	Arg Arg Arg Thr Pro Ser Pro Arg Arg Arg Arg Ser Gln Asn Ser Gln	
111	165 170 175	
113	TTC CAA ACT TGC AAA CAC TTG CCA ACC TCC TGT CCA CCA ACT TGC AAT	576
114	Phe Gln Thr Cys Lys His Leu Pro Thr Ser Cys Pro Pro Thr Cys Asn	
115	180 185 190	
117	GGC TTT CGT TGG ATG TAT CTG CGG CGT TTT ATC ATA TAC CTA TTA GTC	624
118	Gly Phe Arg Trp Met Tyr Leu Arg Arg Phe Ile Ile Tyr Leu Leu Val	
119	195 200 205	
121	CTG CTG CTG TGC CTC ATC TTC TTG TTG GTT CTC CTG GAC TGG AAA GGT	672
122	Leu Leu Leu Cys Leu Ile Phe Leu Leu Val Leu Leu Asp Trp Lys Gly	
123	210 215 220	
125	TTA ATA CCT GTC TGT CCT CTT CAA CCC ACA ACA GAA ACA ACA GTC AAT	720
126	Leu Ile Pro Val Cys Pro Leu Gln Pro Thr Thr Glu Thr Thr Val Asn	
127	225 230 235 240	
130	TGC AGA CAA TGC ACA ATC TCT GCA CAA GAC ATG TAT ACT CCT CCT TAC	768
131	Cys Arg Gln Cys Thr Ile Ser Ala Gln Asp Met Tyr Thr Pro Pro Tyr	
132	245 250 255	
134	TGT TGT TGT TTA AAA CCT ACG GCA GGA AAT TGC ACT TGT TGG CCC ATC	816
135	Cys Cys Cys Leu Lys Pro Thr Ala Gly Asn Cys Thr Cys Trp Pro Ile	
136	260 265 270	
138	CCT TCA TCA TGG GCT TTA GGA AAT TAC CTA TGG GAG TGG GCC TTA GCT	864

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/812,862

DATE: 12/07/2001

TIME: 16:35:08

Input Set : N:\Crf3\RULE60\09812862.txt

Output Set: N:\CRF3\12072001\I812862.raw

```

139 Pro Ser Ser Trp Ala Leu Gly Asn Tyr Leu Trp Glu Trp Ala Leu Ala
140      275      280      285
142 CGT CTC TCT TGG CTC AAT TTA CTA GTG CCC TTG CTT CAA TGG TTA GGA      912
143 Arg Leu Ser Trp Leu Asn Leu Leu Val Pro Leu Leu Gln Trp Leu Gly
144      290      295      300
146 GGA ATT TCC CTC ATT GCG TGG TTT TTG CTT ATA TGG ATG ATT TGG TTT      960
147 Gly Ile Ser Leu Ile Ala Trp Phe Leu Leu Ile Trp Met Ile Trp Phe
148 305      310      315      320
150 TGG GGG CCC GCA CTT CTG AGC ATC TTA CCG CCA TTT ATT CCC ATA TTT      1008
151 Trp Gly Pro Ala Leu Leu Ser Ile Leu Pro Pro Phe Ile Pro Ile Phe
152      325      330      335
154 GTT CTG TTT TTC TTG ATT TGG GTA TAC ATT T GA      1041
155 Val Leu Phe Phe Leu Ile Trp Val Tyr Ile
156      340      345
158 (2) INFORMATION FOR SEQ ID NO: 2:
160 (i) SEQUENCE CHARACTERISTICS:
161 (A) LENGTH: 346 amino acids
162 (B) TYPE: amino acid
163 (D) TOPOLOGY: linear
165 (ii) MOLECULE TYPE: protein
171 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
173 Met Asp Ile Asp Pro Tyr Lys Glu Phe Gly Ser Ser Tyr Gln Leu Leu
174 1      5      10      15
176 Asn Phe Leu Pro Leu Asp Phe Phe Pro Asp Leu Asn Ala Leu Val Asp
177      20      25      30
179 Thr Ala Thr Ala Leu Tyr Glu Glu Glu Leu Thr Gly Arg Glu His Cys
180      35      40      45
182 Ser Pro His His Thr Ala Ile Arg Gln Ala Leu Val Cys Trp Asp Glu
183      50      55      60
185 Leu Thr Lys Leu Ile Ala Trp Met Ser Ser Asn Ile Thr Ser Glu Gln
186 65      70      75      80
188 Val Arg Thr Ile Ile Val Asn His Val Asn Asp Thr Trp Gly Leu Lys
189      85      90      95
191 Val Arg Gln Ser Leu Trp Phe His Leu Ser Cys Leu Thr Phe Gly Gln
192      100      105      110
194 His Thr Val Gln Glu Phe Leu Val Ser Phe Val Val Trp Ile Arg Thr
195      115      120      125
197 Pro Ala Pro Tyr Arg Pro Pro Asn Ala Pro Ile Leu Ser Thr Leu Pro
198      130      135      140
200 Glu His Thr Val Ile Arg Arg Gly Gly Ala Arg Ala Ser Arg Ser Pro
201 145      150      155      160
203 Arg Arg Arg Thr Pro Ser Pro Arg Arg Arg Arg Ser Gln Asn Ser Gln
204      165      170      175
206 Phe Gln Thr Cys Lys His Leu Pro Thr Ser Cys Pro Pro Thr Cys Asn
207      180      185      190
209 Gly Phe Arg Trp Met Tyr Leu Arg Arg Phe Ile Ile Tyr Leu Leu Val
210      195      200      205
212 Leu Leu Leu Cys Leu Ile Phe Leu Leu Val Leu Leu Asp Trp Lys Gly
213      210      215      220

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/812,862

DATE: 12/07/2001

TIME: 16:35:08

Input Set : N:\Crf3\RULE60\09812862.txt

Output Set: N:\CRF3\12072001\I812862.raw

```

215 Leu Ile Pro Val Cys Pro Leu Gln Pro Thr Thr Glu Thr Thr Val Asn
216 225                230                235                240
218 Cys Arg Gln Cys Thr Ile Ser Ala Gln Asp Met Tyr Thr Pro Pro Tyr
219                245                250                255
221 Cys Cys Cys Leu Lys Pro Thr Ala Gly Asn Cys Thr Cys Trp Pro Ile
222                260                265                270
224 Pro Ser Ser Trp Ala Leu Gly Asn Tyr Leu Trp Glu Trp Ala Leu Ala
225                275                280                285
227 Arg Leu Ser Trp Leu Asn Leu Leu Val Pro Leu Leu Gln Trp Leu Gly
228                290                295                300
230 Gly Ile Ser Leu Ile Ala Trp Phe Leu Leu Ile Trp Met Ile Trp Phe
231 305                310                315                320
233 Trp Gly Pro Ala Leu Leu Ser Ile Leu Pro Pro Phe Ile Pro Ile Phe
234                325                330                335
236 Val Leu Phe Phe Leu Ile Trp Val Tyr Ile
237                340                345
239 (2) INFORMATION FOR SEQ ID NO: 3:
241 (i) SEQUENCE CHARACTERISTICS:
242 (A) LENGTH: 1056 base pairs
243 (B) TYPE: nucleic acid
244 (C) STRANDEDNESS: single
245 (D) TOPOLOGY: linear
W--> 247 (ii) MOLECULE TYPE: DNA
253 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
255 CCT CGC AGG ATG GAC ATC GAC CCT TAT AAA GAA TTT GGA GCT ACT GTG      48
256 Pro Arg Arg Met Asp Ile Asp Pro Tyr Lys Glu Phe Gly Ala Thr Val
W--> 257                350                355                360
259 GAG TTA CTC TCG TTT TTG CCT TCT GAC TTC TTT CCT TCA GTA CGA GAT      96
260 Glu Leu Leu Ser Phe Leu Pro Ser Asp Phe Phe Pro Ser Val Arg Asp
W--> 261                365                370                375
263 CTT CTA GAT ACC GCC TCA GCT CTG TAT CGG GAA GCC TTA GAG TCT CCT      144
264 Leu Leu Asp Thr Ala Ser Ala Leu Tyr Arg Glu Ala Leu Glu Ser Pro
W--> 265                380                385                390
267 GAG CAT TGT TCA CCT CAC CAT ACT GCA CTC AGG CAA GCA ATT CTT TGC      192
268 Glu His Cys Ser Pro His His Thr Ala Leu Arg Gln Ala Ile Leu Cys
W--> 269 395                400                405                410
271 TGG GGG GAA CTA ATG ACT CTA GCT ACC TGG GTG GGT GTT AAT TTG GAA      240
272 Trp Gly Glu Leu Met Thr Leu Ala Thr Trp Val Gly Val Asn Leu Glu
W--> 273                415                420                425
275 GAT CCA GCG TCT AGA GAC CTA GTA GTC AGT TAT GTC AAC ACT AAT ATG      288
276 Asp Pro Ala Ser Arg Asp Leu Val Val Ser Tyr Val Asn Thr Asn Met
W--> 277                430                435                440
279 GGC CTA AAG TTC AGG CAA CTC TTG TGG TTT CAC ATT TCT TGT CTC ACT      336
280 Gly Leu Lys Phe Arg Gln Leu Leu Trp Phe His Ile Ser Cys Leu Thr
W--> 281                445                450                455
283 TTT GGA AGA GAA ACA GTT ATA GAG TAT TTG GTG TCT TTC GGA GTG TGG      384
284 Phe Gly Arg Glu Thr Val Ile Glu Tyr Leu Val Ser Phe Gly Val Trp
W--> 285                460                465                470
287 ATT CGC ACT CCT CCA GCT TAT AGA CCA CCA AAT GCC CCT ATC CTA TCA      432

```

RAW SEQUENCE LISTING

DATE: 12/07/2001

PATENT APPLICATION: US/09/812,862

TIME: 16:35:08

Input Set : N:\Crif3\RULE60\09812862.txt

Output Set: N:\CRF3\12072001\I812862.raw

```

288 Ile Arg Thr Pro Pro Ala Tyr Arg Pro Pro Asn Ala Pro Ile Leu Ser
W--> 289 475          480          485          490
291 ACA CTT CCG GAA CAT ACA GTC ATT AGA AGA GGA GGT GCA AGA GCT TCT      480
292 Thr Leu Pro Glu His Thr Val Ile Arg Arg Gly Gly Ala Arg Ala Ser
W--> 293          495          500          505
295 AGG TCC CCC AGA AGA CGC ACT CCC TCT CCT CGC AGG AGA AGA TCC CAA      528
296 Arg Ser Pro Arg Arg Arg Thr Pro Ser Pro Arg Arg Arg Arg Ser Gln
W--> 297          510          515          520
299 AAT TCG CAG TCC CCA ACC TCC AAT CAC TCA CCA ACC TCT TGT CCT CCA      576
300 Asn Ser Gln Ser Pro Thr Ser Asn His Ser Pro Thr Ser Cys Pro Pro
W--> 301          525          530          535
303 ACT TGT CCT GGT TAT CGC TGG ATG TGT CTG CGG CGT TTT ATC ATC TTC      624
304 Thr Cys Pro Gly Tyr Arg Trp Met Cys Leu Arg Arg Phe Ile Ile Phe
W--> 305          540          545          550
307 CTC TTC ATC CTG CTG CTA TGC CTC ATC TTC TTG TTG GTT CTT CTG GAC      672
308 Leu Phe Ile Leu Leu Leu Cys Leu Ile Phe Leu Leu Val Leu Leu Asp
W--> 309 555          560          565          570
311 TAT CAA GGT ATG TTG CCC GTT TGT CCT CTA ATT CCA GGA TCC TCA ACA      720
312 Tyr Gln Gly Met Leu Pro Val Cys Pro Leu Ile Pro Gly Ser Ser Thr
W--> 313          575          580          585
315 ACC AGC ACG GGA CCA TGC CGG ACC TGC ATG ACT ACT GCT CAA GGA ACC      768
316 Thr Ser Thr Gly Pro Cys Arg Thr Cys Met Thr Thr Ala Gln Gly Thr
W--> 317          590          595          600
319 TCT ATG TAT CCC TCC TGT TGC TGT ACC AAA CCT TCG GAC GGA AAT TGC      816
320 Ser Met Tyr Pro Ser Cys Cys Cys Thr Lys Pro Ser Asp Gly Asn Cys
W--> 321          605          610          615
323 ACC TGT ATT CCC ATC CCA TCA TCC TGG GCT TTC GGA AAA TTC CTA TGG      864
324 Thr Cys Ile Pro Ile Pro Ser Ser Trp Ala Phe Gly Lys Phe Leu Trp
W--> 325          620          625          630
327 GAG TGG GCC TCA GCC CGT TTC TCC TGG CTC AGT TTA CTA GTG CCA TTT      912
328 Glu Trp Ala Ser Ala Arg Phe Ser Trp Leu Ser Leu Leu Val Pro Phe
W--> 329 635          640          645          650
331 GTT CAG TGG TTC GTA GGG CTT TCC CCC ACT GTT TGG CTT TCA GTT ATA      960
332 Val Gln Trp Phe Val Gly Leu Ser Pro Thr Val Trp Leu Ser Val Ile
W--> 333          655          660          665
335 TGG ATG ATG TGG TAT TGG GGG CCA AGT CTG TAC AGC ATC TTG AGT CCC      1008
336 Trp Met Met Trp Tyr Trp Gly Pro Ser Leu Tyr Ser Ile Leu Ser Pro
W--> 337          670          675          680
339 TTT TTA CCG CTG TTA CCA ATT TTC TTT TGT CTT TGG GTA TAC ATT T      1054
340 Phe Leu Pro Leu Leu Pro Ile Phe Phe Cys Leu Trp Val Tyr Ile
W--> 341          685          690          695
343 AA      1056
345 (2) INFORMATION FOR SEQ ID NO: 4:
347 (i) SEQUENCE CHARACTERISTICS:
348 (A) LENGTH: 351 amino acids
349 (B) TYPE: amino acid
350 (D) TOPOLOGY: linear
352 (ii) MOLECULE TYPE: protein
358 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/812,862

DATE: 12/07/2001

TIME: 16:35:09

Input Set : N:\Crf3\RULE60\09812862.txt

Output Set: N:\CRF3\12072001\I812862.raw

L:29 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]
L:30 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]
L:59 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=1
L:247 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=3
L:257 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:261 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:265 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:269 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:273 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:277 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:281 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:285 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:289 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:293 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:297 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:301 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:305 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:309 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:313 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:317 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:321 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:325 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:329 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:333 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:337 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:341 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:434 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=5
L:445 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:449 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:453 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:457 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:461 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:465 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:469 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:473 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:477 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:481 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:485 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:489 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:493 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:497 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:501 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:505 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:509 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:513 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:517 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:521 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:525 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5